

Tenorth, Heinz-Elmar

## **Wer hat Angst vor den "Evaluationskommissaren"? Evaluation und Qualitätssteigerung der erziehungswissenschaftlichen Forschung**

*Zeitschrift für Pädagogik 50 (2004) 1, S. 88-98*



Quellenangabe/ Reference:

Tenorth, Heinz-Elmar: Wer hat Angst vor den "Evaluationskommissaren"? Evaluation und Qualitätssteigerung der erziehungswissenschaftlichen Forschung - In: Zeitschrift für Pädagogik 50 (2004) 1, S. 88-98 - URN: urn:nbn:de:0111-opus-47999 - DOI: 10.25656/01:4799

<https://nbn-resolving.org/urn:nbn:de:0111-opus-47999>

<https://doi.org/10.25656/01:4799>

in Kooperation mit / in cooperation with:

# **BELTZ**

<http://www.beltz.de>

### **Nutzungsbedingungen**

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### **Terms of use**

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

### **Kontakt / Contact:**

**peDOCS**  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Mitglied der

  
Leibniz-Gemeinschaft

## Inhaltsverzeichnis

### *Thementeil: Bildung über die Lebenszeit*

<i>Reinhard Fatke</i>	
Bildung über die Lebenszeit – Einführung in den Thementeil .....	1
<i>Aleida Assmann</i>	
Der väterliche Bücherschrank – Über Vergangenheit und Zukunft der Bildung .....	5
<i>Rita Casale/Jürgen Oelkers/Daniel Tröhler</i>	
Lebenslanges Lernen in historischer Perspektive – Drei Beispiele für ein altes Konzept .....	21
<i>Ludwig Liegle/Kurt Lüscher</i>	
Das Konzept des „Generationenlernens“ .....	38
<i>Helmut Fend/Fred Berger/Urs Grob</i>	
Langzeitwirkungen von Bildungserfahrungen am Beispiel von Lesen und Computer Literacy – Ergebnisse der LIFE-Studie (Lebensverläufe von der späten Kindheit ins frühe Erwachsenenalter) .....	56
 <i>Diskussion 1: Evaluation der Erziehungswissenschaft</i>	
<i>Hans Merrens</i>	
Evaluation in der Erziehungswissenschaft – eine neue Herausforderung? .....	77
<i>Heinz-Elmar Tenorth</i>	
Wer hat Angst vor den „Evaluationskommissaren“? Evaluation und Qualitätssteigerung der erziehungswissenschaftlichen Forschung .....	88
<i>Frieda Heyting</i>	
Beobachtungen zur internationalen Anschlussfähigkeit der Allgemeinen Erziehungswissenschaft in Deutschland .....	99

## *Diskussion 2: „Neue Kindheit“ in der Diskussion*

### *Horst Rumpf*

Für einen entdramatisierenden Umgang mit der „neuen Kindheit“?

Zur Abwehr pädagogischer Horrorvisionen.

Eine Erwiderung auf Jürgen Oelkers ..... 112

### *Jürgen Oelkers*

Eine Erwiderung auf Horst Rumpf ..... 116

## *Sammelbesprechung*

### *Hans-Joachim von Olberg*

Didaktik auf dem Wege zur Vermittlungswissenschaft?

Eine Sammelbesprechung neuerer Veröffentlichungen ..... 119

## *Besprechungen*

### *Joachim Kersten*

Rudolf Leiprecht: Alltagsrassismus. Eine Untersuchung bei Jugendlichen in  
Deutschland und den Niederlanden

Benno Hafener/Mechtild M. Jansen: Rechte Cliques. Alltag einer neuen  
Jugendkultur

Edith Wölfl: Gewaltbereite Jungen – was kann Erziehung leisten? Anregungen  
für eine gender-orientierte Pädagogik ..... 132

### *Gisela Miller-Kipp*

Karl Heinz Jahnke: Jugend unter der NS-Diktatur 1933–1945. Eine Dokumentation

Michael Buddrus: Totale Erziehung für den totalen Krieg. Hitler-Jugend und  
nationalsozialistische Jugendpolitik. 2 Bde.

Hermann Langer: „Im gleichen Schritt und Tritt“. Die Geschichte der  
Hitler-Jugend in Mecklenburg von den Anfängen bis 1945 ..... 138

### *Alois Suter*

Robert Döpp: Jenaplan-Pädagogik im Nationalsozialismus. Ein Beitrag zum

Ende der Eindeutigkeit ..... 142

## *Dokumentation*

Pädagogische Neuerscheinungen ..... 146

Heinz-Elmar Tenorth

## Wer hat Angst vor den „Evaluationskommissaren“?

*Evaluation und Qualitätssteigerung der erziehungswissenschaftlichen Forschung*

„Vom Wiegen wird das Schwein nicht fett.“  
(Hartmut Holzapfel)

**Zusammenfassung:** Die Leitfrage des Beitrags lautet: Kann durch Evaluation ein Beitrag zur Entwicklung und Verbesserung der Forschung in der Erziehungswissenschaft geleistet werden? Ausgehend von dem damit angesprochenen Zusammenhang von Diagnose und Konstruktion werden beim Blick auf die bisherige Evaluationspraxis einige Defizite formuliert. Generell wird jedoch den bei der Evaluation von wissenschaftlichen Disziplinen üblicherweise angewandten Verfahren und Indikatoren auch bei der Evaluation der Erziehungswissenschaft durchaus eine Aussagekraft zugebilligt, wobei zugleich auf die Notwendigkeit von besonderen, disziplintypischen Spezifikationen – etwa in der Lehrerbildung – hingewiesen wird. Über die Forschung nicht nur evaluierende, sondern aufgrund bestimmter praktischer Schlussfolgerungen auch fördernde Wirkung von Evaluationen gibt es für die Erziehungswissenschaft allerdings noch keine empirisch gestützten Erkenntnisse.

### 1. Die Situation

Das Bild ist vertraut, dennoch ist die Wahrnehmung der im Prinzip vertrauten Situation immer wieder überraschend: Vertraut ist die Situation, dass die Leistungen von Wissenschaften bewertet werden, vor allem in einer Disziplin wie der Erziehungswissenschaft, deren Geschichte in Deutschland und weltweit sich seit mehr als 200 Jahren als ein Prozess kritischer Selbst- und Fremdbeobachtung beschreiben und in der Kontinuität scharfer evaluativer Urteile, interner wie externer Provenienz, über ihren Status als Wissenschaft und über ihre Leistungen in der Forschung dokumentieren lässt (Lagemann 2000; Tenorth 2003).

Obwohl die Disziplin deshalb im Umgang mit Selbst- und Fremdbeobachtung eigentlich erfahren und geübt sein sollte und vielleicht gelassen oder sogar im Modus des Lernens hätte reagieren können (wie in Keiner 2001), führen aktuelle Evaluationsverfahren, wie es sie in Bayern, Niedersachsen oder Baden-Württemberg gibt oder jüngst gegeben hat, erneut zu erstaunlichen Reaktionen. Offenbar brodet es in der Tiefe der Disziplin: Ein bayerischer Ordinarius, von Evaluationsverfahren und -befunden schmerzlich betroffen (Hörmann 2002), macht der Öffentlichkeit seinen ganzen Unmut in einer Suada von Vorwürfen bekannt, die aber weder den Adressaten noch die Anlässe seiner Schelte richtig zuschreiben und sich zudem auf den Text einer Evaluationskommission und begleitende Dokumente der Abnehmer beziehen, die der Öffentlichkeit gar nicht zugänglich sind (und wenigstens diesen Missstand hat das apologetische Pamphlet mit guten Gründen in Erinnerung gerufen). Mag man solche Reaktionen noch als vereinzelte und verständliche Abwehr eines Kritisierten charakterisieren, so ist schon eher irritierend, dass ein sehr renommierter, wenn auch für kritische Urteile bekannter

Kollege aus der Allgemeinen Erziehungswissenschaft die Sprache des Kalten Krieges nicht scheut, wenn er über Evaluationsverfahren schreibt. Man liest mit Erstaunen, dass hier „Evaluationskommissare“ am Werk seien, rekrutiert aus den „Großmeistern“ der Disziplin (Prange 2003, S. 756)<sup>1</sup>, die in der Erziehungswissenschaft diejenigen – wenigen, braven? – rüde bewerten, die noch „tatsächlich etwas erforschen oder in der Lehre aufgehen“ (ebd.). Man soll das nicht als Einzelfall abtun; selbst Vorstandsmitglieder der DGfE (Merkens/Tippelt 2002) und eine eigene Veranstaltung der GEW haben fast erregt auf den Vortrag eines amerikanischen Kollegen reagiert, der zwar kein gerechtes Bild der gesamten Erziehungswissenschaft zeichnet, aber doch mit guten Gründen, wenn auch mit Verve festgehalten hat, dass in der deutschen empirischen Bildungsforschung thematisch erhebliche Lücken existieren, besonders bei bildungspolitisch brisanten Fragen wie z.B. der Effektivität und Wirkung der Lehrerbildung, und dass die Disziplin zugleich erstaunliche Präferenzen auch gegenüber der Bildungspolitik zeige, nämlich ein eher politikritisches und distantes als an Beratung und Unterstützung interessiertes Verhalten (Weiler 2003).

Aus der Distanz wird man solche Reaktionen als Teil der inzwischen auch international gewachsenen und deutlicher werdenden Selbstthematisierung der Erziehungswissenschaft und der Bildungsforschung einordnen, denn unverkennbar wird diese Disziplin (wie andere) nicht allein in Deutschland neu vermessen und kritisch beurteilt. Die kritischen Kommentare z.B. der Innsbrucker Kollegen auf die externe Evaluation ihres Faches, das zeigte die web-site des Instituts, unterschieden sich 2002 in ihrer Heftigkeit kaum von deutschen Reaktionen. Aber auch hier lag eine Beschreibung der österreichischen Erziehungswissenschaft bereits vor, die kaum weniger kritisch den Zustand der gesamten Disziplin gekennzeichnet hatte (Brezinka 2000) – man hätte also wissen können, was Evaluationen zutage fördern.

„Krisen“-diagnosen, mit denen frühere Zustandsbeschreibungen der Erziehungswissenschaft häufig identisch waren, gibt es nicht nur erneut für die deutsche Erziehungswissenschaft (Brezinka 2003), sondern im Kontext ausführlicher Selbstevaluationen auch für die internationale Lage der Disziplin (Badley 2003), auch hier nicht zufällig in der Bekräftigung der eigenen Position als Ausweg aus der Krise. Die Evaluation ganzer Disziplinen im nationalen Rahmen hat inzwischen selbst Tradition, in den Niederlanden z.B. für die Wissenschaft des Landes, aber auch für die Erziehungswissenschaft (Rang 2001) und mit vorbildhafter Wirkung für ganz Europa (van der Meulen 2003). Vergleichbare Stetigkeit und Systematik hat die Forschungsevaluation in Großbritannien gewonnen, im Allgemeinen (Martin 2003) und in einer kritischen Evaluation der erziehungswissenschaftlichen Forschung (OECD 2002). Umfassende Selbstbeschreibungen finden sich z.B. auch für die Schweiz (Hofstetter/Schneuwly 2001), auch hier vor dem Hintergrund systematischer Beobachtung der Erziehungswissenschaft im internationalen Kontext (Hofstetter/Schneuwly 1998, 2002).

1 An anderer Stelle (Prange i. Druck) befürchtet Klaus Prange sogar, dass bald „Proskriptionslisten“ ausgehängt würden, um unwillige oder leistungsschwache Kollegen öffentlich zu brandmarken – und dann auszuschalten, wie weiland im alten Rom, wo diese Praxis ihren Ursprung hat?

Die Karriere solcher Analysen erklärt sich selbstverständlich nicht allein aus den Karriereinteressen junger Wissenschaftler, die sich für Wissenschaftsforschung interessieren, sie ist vielmehr integraler Teil der kritischen Bewertung von Wissenschaften und von Hochschulsystemen überhaupt, wie sie international seit mehr als zwei Jahrzehnten zu beobachten sind. Dabei sind Kriterien der Rationalität des Mitteleinsatzes und der internationalen Offenheit der Bildungsmärkte leitend, aber auch die Erwartung, dass aus besserer Bildungsforschung schließlich auch eine Qualitätssteigerung im Bildungswesen resultieren könnte.

Damit ist der Punkt erreicht, an dem Evaluationen ihre wirkliche Relevanz gewinnen, in dem Anspruch nämlich, nicht allein die Wissenschaftler über sich selbst zu belehren (das wäre kaum so umfassend nötig gewesen, weil in der scientific community Differenzen in der Qualität der Forschung durchaus bewusst waren, auch wenn man sie öffentlich nicht so deutlich kommuniziert hat), sondern mit der Verbesserung der Wissenschaft auch die Qualität von Bildungssystemen und -prozessen zu steigern. Dieser Zusammenhang wird jedoch selten diskutiert, allenfalls skeptisch für die Effekte der Evaluation der Lehrerbildung (von Prondczynsky 2001). Eine Diskussion dieses Anspruchs ist indes notwendig, damit Evaluation nicht zu einem selbstgenügsamen Betrieb verkommt.

Die Diskussion sollte sich vor allem mit der Frage beschäftigen, ob in der Alltätlichkeit der Arbeit von Kommissionen und „Evaluationskommissaren“ mehr als der Betrieb der Evaluation befördert wird, oder ob sich tatsächlich Indizien dafür finden lassen, dass diese Arbeit den Zusammenhang von Evaluation und Qualitätssteigerung hinreichend in den Blick nimmt und im Ergebnis sogar befördert, zumindest für die Forschung, wenn man schon nicht das ganze Bildungssystem verändern und verbessern will. Weil die letzte Frage im Kontext von PISA schon diskutiert worden ist, konzentrieren sich die folgenden Überlegungen auf die Möglichkeiten der Entwicklung und Verbesserung von Forschung im Kontext von Evaluationsprozessen.

## **2. Evaluation**

Was immer Evaluation bedeutet, sie stellt „ein Bündel von Verfahren“ dar, mit dem „eine wissenschaftsgestützte Beschreibung und Analysen mit praktisch folgenreichen Bewertungen“ verbunden werden sollen (Hornbostel/Keiner 2002, S. 634). In Evaluationsverfahren zur Forschung wird also ein Urteil über die einschlägige Praxis einer Disziplin in der Absicht formuliert, Leistungen und Defizite, Stärken und Schwächen angesichts gegebener Kriterien, bzw. in der Relation, d.h. im Vergleich mit anderen Fächern, oder innerhalb der jeweiligen Disziplin mit anderen Teildisziplinen oder Personen festzustellen. Solche Leistungsbehauptungen oder Defizitbefunde werden vor dem Horizont von Erwartungen vorgetragen, künftig Stärken zu belohnen, Schwächen abzustellen und die Qualität der Forschung zu steigern. Wie immer man über Evaluation denkt, allein diese Ziele machen die Anstrengungen sinnvoll und die in der Bewertung von Personen liegende Zudringlichkeit gegenüber Kollegen überhaupt erst legitim.

Lässt man zunächst die Frage außer Acht, wie die Anschlussstrategien aussehen können, dann muss man zuerst jedes Evaluationsverfahren darauf hin prüfen, ob es über die angemessene Theorie und über die geeigneten Instrumente für die intendierte Statusdiagnose verfügt. Dabei wird man von den Instrumenten nicht allein die Qualität erwarten, die in der Wissenschaftsforschung wie in der gesamten empirischen Sozialforschung bei der Beschreibung und Analyse von Sachverhalten üblich ist, also die Gütekriterien der Reliabilität, Validität und Objektivität, sondern auch prognostische Fähigkeiten. Erst wenn die Grundlagen der Arbeit und die Ursachen für Leistungen und Defizite des Faches diagnostisch eindeutig bezeichnet sind, können auch Verbesserungsanstrengungen sinnvoll ansetzen. Evaluation stellt nicht nur selbst einen Prozess dar, der nach Kriterien der Forschung bewertet werden kann, ja muss, er wiederholt in der Verbindung von Diagnose, Wissen und Konstruktion auch generell ein altes Motiv der empirischen Forschung: *savoir pour prévoir!*, das ist die Logik der Evaluation.

Prüft man vor diesem Hintergrund die Evaluationspraxis und -reflexion in der Erziehungswissenschaft und für die Disziplin, dann wird man schwerlich behaupten können, dass die Dimensionen dieses Anspruchs, Diagnose und Konstruktion, bereits hinreichend diskutiert sind. Sicherlich, man wird nicht mehr behaupten, dass die Qualität wissenschaftlicher Arbeit sich generell der Evaluation entzieht oder dass Forschung in einzelnen Disziplinen nicht vergleichend messbar sei. Solche Aktivitäten sind nicht allein durch die abendländische philosophische Tradition wissenschaftlicher Selbstreflexion, sondern in jüngerer Zeit auch durch die Praxis der Wissenschaftsforschung hinreichend legitimiert worden. Evaluationsverfahren profitieren von dieser Tradition und haben die einschlägige Praxis durch die Entwicklung von Standardindikatoren selbst bereichert.

Die in der Regel benutzten Indikatoren bilden dabei Dimensionen von Wissenschaft ab, die wirklich für deren Praxis zentral sind: Formen der Kommunikation von Forschung, wie sie in Zitationsindizes und Publikationsfrequenzen und -orten diskutierbar werden, sowie die Bereitschaft zur Innovation, wie sie außer über Publikationen in begutachteten, durch Drittmittel unterstützten Forschungsprogrammen dokumentiert wird. Beide Indikatorsysteme, für Publikationen wie für Drittmiteinsatz, machen die Qualität von Forschung sichtbar, und sie tun das im Wesentlichen dadurch, dass sie die kollegiale Anerkennung und Sichtbarkeit der Forschungspraxis sowie ihr Innovationspotenzial zum Maßstab für Qualität machen. Forschung wird in der Kommunikation einer Disziplin erzeugt und bewertet, das ist die wissenschaftstheoretische Prämisse, die zugleich Prozess und Produkt der Arbeit verbindet.

Diese Indikatoren, obwohl in ihren Datengrundlagen immer verbesserungsfähig (Weishaupt/Preuschoff 2002), sind für die Analyse also gut geeignet. Sie leiden auch nicht an dem häufig zugeschriebenen Fehler, allein ein Normalmodell von Forschung sichtbar zu machen; sie sind vielmehr in den Merkmalsausprägungen sensibel für die Eigenheiten und Besonderheiten von Disziplinen: Mit Zitationsindizes konnten in der jüngeren Wissenschaftsforschung nicht nur nationale Kulturen innerhalb von Erziehungswissenschaft und Bildungsforschung herausgearbeitet werden (Keiner/Schriewer 2000), sondern auch unterschiedliche Positionen der Erziehungswissenschaft und Indi-

zien für ihre Binnenorientierung oder Offenheit innerhalb der Sozialwissenschaften. Aus der vergleichenden Analyse der Drittmittelforschung wiederum ergeben sich nicht nur Hinweise auf Defizite im Antragsvolumen oder in den gesuchten Drittmittelgebern, z.B. im Zugang zur DFG (Nießen 2001), sondern auch unerwartete Gemeinsamkeiten und Unterschiede, wenn man z.B. erfährt (Hornbostel/Keiner 2002, bes. S. 646ff.), dass die Erziehungswissenschaft in dieser Beziehung eher den Rechtswissenschaften vergleichbar ist als den Sozialwissenschaften (ohne dass sie sich das immer eingesteht).

Die Irritation von Selbstbildern, die solche Praxis von Evaluation und Wissenschaftsforschung ergibt, erweist sich auch für die theoretischen Implikationen der Evaluation selbst. So ist z.B. Drittmittelforschung in ihrem Status nicht immer von gleicher Qualität, zwingend kommunikationsfördernd in der Gemeinschaft der Forscher oder stets praxisrelevant. Schon länger ist bekannt, dass die in der Erziehungswissenschaft nicht seltenen Projekte, die sich aus Etats öffentlich-staatlicher Geldgeber, z.B. der Ministerien für Bildung oder Soziales, speisen, weniger zu Publikationen führen als Forschung im Kontext der DFG-Förderung (Baumert/Roeder 1990). Drittmittel bedürfen also selbst der Bewertung, z.B. nach den Verfahren der Begutachtung oder nach der Zugänglichkeit und Verfügbarkeit am Wissenschaftsmarkt und der damit gegebenen konjunkturellen Situation. Sie hat z.B. in den letzten Jahren den Zugang zu Mitteln im Kontext neuer Medien bzw. von IT-Strategien und -programmen im Bildungsbereich relativ einfach und lukrativ eröffnet; „Forschung“ war mit diesen Programmen nicht immer verbunden, und man wird deshalb evaluativ prüfen müssen, ob die Entwicklungsarbeit wirklich hinreichend kompensiert, was an Forschungsqualität fehlte.

Insgesamt: Die methodisch und theoretisch reflektierte Nutzung der Standardindikatoren eröffnet sehr viel mehr an Einsichten über die Erziehungswissenschaft, als es sich die rasche Kritik an solchen Standardinstrumenten eingesteht, ja man erkennt in der Kritik der Indikatoren eher die Abwehr externer Prüfung und interner Rechenschaftslegung als wirklich ein begründetes Verfahren der Kritik von Evaluation. Es ist deshalb auch nicht Indiz einer Verlegenheit, sondern eine sachlich angemessene Entscheidung, wenn die DGfE ihre „Datenreports“ im Blick auf die eigene Forschung auch über solche Standardindikatoren organisiert und empirisch absichert (Weishaupt/Merkens 2000; Merkens/Weishaupt/Zedler 2000; Merkens/Dreyer 2002; Weishaupt 2002).

Aus der Gegenüberstellung von Forschung und Entwicklung ergibt sich andererseits schon ein Verweis auf ein wichtiges Anschlussproblem. Neben der angemessenen Verortung eines Faches als Wissenschaft im Forschungsprozess, wie sie sich aus der Praxis der Standardindikatoren ergeben kann, ist die disziplinspezifische Aussagekraft selbstverständlich ein eigenes Thema. In der frühen Wissenschaftsforschung bereits aufgenommen, erweist es sich in Evaluationsverfahren als durchaus nicht einfach, z.B. den spezifischen „praxisbezogenen“ Forschungstypus, den die Erziehungswissenschaft für sich reklamiert, in ihrer Praxis auch zu identifizieren (Macke 1990 u.ö.). Etwas überrascht muss man immer neu feststellen, dass in der Erziehungswissenschaft auch in dieser Hinsicht zwischen der wissenschaftstheoretischen Propaganda und den starken Behauptungen der Eigenständigkeit in der Forschung einerseits und der Praxis der Arbeit andererseits große Differenzen bestehen. „Praxisbezogene Wissenschaft“, damit wird



eher die Stilisierung eines Programms als eine angemessene Selbstbeschreibung ihrer Praxis gegeben.

Man hatte also nicht viel verpasst, wenn man den Alltag der Forschung nur mit den allgemein üblichen Indikatoren und Kriterien für Forschung betrachtete. Aber selbstverständlich bleiben trotz solcher Befunde weitere Anstrengungen notwendig, das spezifische Forschungsprofil des Faches zu identifizieren. Ausgehend von ihren Selbstbildern kann die Erziehungswissenschaft dann z.B. auch als „Berufswissenschaft“ – für Lehrer oder auch für die Soziale Arbeit – verstanden werden und die Frage ist legitim, ob sich von daher ein spezifischer Forschungstypus identifizieren lässt. In der Sozialen Arbeit ist daraus ein bis heute ungeschlichteter Grundlagenstreit entstanden, genährt auch aus der Abgrenzung von Fachhochschulen und Universitäten und aus der wechselseitigen Zuschreibung von Relevanz und Irrelevanz der eigenen Arbeit. In Evaluationsverfahren ist die Fokussierung auf die berufswissenschaftliche Seite der Forschung bisher erst einmal versucht worden (Wissenschaftliche Kommission Niedersachsen 2002)<sup>2</sup>. Unter dem Begriff der „Berufswissenschaften der Lehrerbildung“ wurden hier diejenigen Disziplinen zusammengefasst, die Wissen bereitstellen und Probleme erforschen, die einerseits „für die operative Dimension“ der Lehrertätigkeit von Bedeutung sind, also „für die Gestaltung der pädagogischen Interaktion und Kommunikation, für die didaktische Rekonstruktion fachlicher Inhalte, für die Diagnose von Lernmöglichkeiten und -schwierigkeiten, für die Orientierung in der Organisation und im professionellen und sozialen Kontext, für die Reflexion von Erwartungen und Zumutungen an das Bildungssystem“; dazu rechnen aber auch Disziplinen, die als „ihre Referenz nicht allein die lehrenden Personen, sondern die Funktionsprobleme des Bildungssystems insgesamt“ suchen (ebd. S. 12f.). Einschlüssig sind damit neben der Erziehungswissenschaft und ihren schulpädagogischen Subdisziplinen auch die Pädagogische Psychologie, die Soziologie der Erziehung sowie die Bereichs- und Fachdidaktiken. Forschung in diesem berufswissenschaftlichen Sinne ist deshalb auch nicht allein in den Standardindikatoren abbildbar, denn sie bezeichnen nur notwendige, nicht hinreichende Möglichkeiten der Identifikation und Qualifizierung, sie bedarf der thematischen Spezifikation (und dann war das Ergebnis eher durch große Heterogenität der Forschungsleistungen, auch durch große Defizite, als durch ein wirklich befriedigendes Gesamtbild gekennzeichnet).

In solcher thematisch orientierten Fragestellung zeigen sich sehr rasch die Grenzen der meist genutzten Instrumente. Maße für Kommunikation oder Drittmiteleinwerbung sind zunächst neutral gegenüber inhaltlichen Erwartungen, und sie sind auch nicht geeignet, zukunftsfähige Themen zu identifizieren oder Desiderata qualitativ auszuzeichnen. Das sind aber bedeutsame Fragen für den Status und die Perspektive eines Faches, und Antworten auf diese Fragen sind legitime Erwartungen an Evaluationsverfahren. Sie können deshalb auch sinnvoll gar nicht veranstaltet werden, ohne ein systematisches Bild der jeweiligen Disziplin, ihrer Binnenstruktur und ihrer Aufgaben.

2 Die Evaluationskommission habe ich selbst geleitet, sodass ich mich qualitätsbezogener Argumente über ihre Arbeit enthalte; es kommt mir nur darauf an, ein Thema in Erinnerung zu rufen, das für die behauptete Spezifik der erziehungswissenschaftlichen Forschungsarbeit von Bedeutung ist.

Defizitdiagnosen und Leistungsbehauptungen können dann nicht allein auf Standardindikatoren beruhen; in den systematischen Fragen beruhen sie auf den Urteilen der peers, die in den Evaluationskommissionen mitwirken. Vor diesem Hintergrund kann man die Sorge über die Rolle der „Großmeister“ verstehen und die „Kommissar“-Metaphorik als vorbeugende Kritik an einseitigen Urteilen und verzerrten Diagnosen und Prognosen in ihrem Problemgehalt anerkennen. Aber die Lösung des Problems kann nicht in der Enthaltensamkeit der Urteile bestehen oder in Mehrheitsabstimmungen über die Qualität der Forschung, sondern nur in der offenen und offensiven Kommunikation über die evaluativen Befunde. Die dann notwendigen Kontroversen über Präferenzen und Optionen der Fachentwicklung befördern die Entwicklung des Faches freilich erst dann, wenn sie in prüfbaren Daten ihr Fundament haben.

### **3. Qualitätssteigerung**

Evaluation bleibt ein Glasperlen- und Machtspiel der Evaluateure und ihrer Auftraggeber, wenn sie nicht zu Konsequenzen führt. Sie bliebe auch dann kaum mehr als eine Irritation der von ihr Evaluierten, wenn sie nicht in solche Konsequenzen mündete, die wirklich zielbezogen zur Steigerung der Qualität von Forschung führen. Wenn der frühere hessische Kultusminister Hartmut Holzapfel – bezogen auf die assessment-Studien über Schulen und Bildungssysteme – das inzwischen inkriminierte Diktum fällt: „Vom Wiegen wird das Schwein nicht fett“, dann hat er mit guten Gründen den notwendig herzustellen, aber nicht selbstverständlichen Zusammenhang von Evaluation und Qualitätssteigerung in Erinnerung gerufen. Diese Verknüpfung muss auch angesichts von Forschungsevaluationen gefordert werden. Dabei weiß selbstverständlich auch Holzapfel, vielleicht – um im Bilde zu bleiben – aus der Schweinemast, dass auch das Wiegen notwendig ist, um die Angemessenheit und Effizienz der eigenen Maßnahmen zu kontrollieren und gegebenenfalls die eigenen Strategien zu revidieren. Aber er fordert zu Recht mehr, Fördermaßnahmen z.B., um die Qualität zu steigern.

Was muss man aber der Wissenschaft zuführen, quasi als nutrimentum spiritus, damit ihre Qualität wächst? Gilt für alle Wissenschaften das gleiche Programm oder bedürfen besondere Disziplinen auch besonderer Zuwendung? Mir scheint das der zentrale Punkt in der Diskussion über Evaluation zu sein, und zwar der zentrale offene Punkt, in Details wie in den Wirkungen weitgehend ungeklärt. Sieht man einmal von der Tatsache ab, dass manche Evaluationsvorhaben schon die Evaluation selbst für ein zentrales Instrument disziplinärer Verbesserung halten, weil dadurch das Bewusstsein für die disziplinäre Identität geschärft werde (van der Meulen 2003, S. 279); die Rationalität des Verfahrens und die prozessbegleitend quasi automatisch eintretenden Effekte reichen allein nicht aus. Zwar verändert sich Wissenschaft im Medium der Öffentlichkeit, aber Abwehr von Kritik ist auch Veränderung, und angesichts der Kontinuität der Krisendiagnosen in der Erziehungswissenschaft wird man zuverlässig darüber belehrt, dass die öffentliche Präsenz von Kritik keineswegs so viel Veränderungsdruck erzeugt, dass eine Disziplin ihren Idiosynkrasien abschwört. Auch die allein innerfachliche Kommunika-

tion bringt kaum radikalen Wandel, sonst hätte der relativ hohe Konsens, der in allen modernen Wissenschaften über Qualität und Exzellenz existiert, die Konsequenz gehabt, auch Schwächen des jeweiligen Faches nicht zu tolerieren. Letzteres war bekanntlich nicht der Fall (so wenig wie die Selbstkontrolle in den Naturwissenschaften z.B. zuverlässig die Betrugsmanöver verhindert hat, die in der jüngsten Zeit publik geworden sind).

Die offene Frage bleibt also, wie man Qualitätssteigerung befördern kann, wenn die Alltagsmechanismen wissenschaftlicher Arbeit, von revolutionären Umbrüchen abgesehen, eher Kontinuität und die wechselseitige Bestätigung in den eigenen Revieren und Traditionen bekräftigen. Problematisch und schwierig veränderbar ist diese Situation wegen eines systematischen Konflikts, der sich beim Versuch der Steuerung von Wissenschaften im Forschungsprozess ergibt, nämlich der Konflikt zwischen der Freiheit von Forschung und Wissenschaft einerseits und der Steuerungsabsicht andererseits.<sup>3</sup> Die Strategien der Steuerung können in Staaten, in denen die Freiheit der Forschung anerkannt ist, deshalb auch nicht auf dem Prinzip unmittelbarer Eingriffe beruhen, sondern nur indirekt Einfluss suchen: durch Empfehlungen versuchen es Evaluationskommissionen; orientiert an Normen und daraus abgeleiteten Programmen und Vorgaben arbeiten andere, z.B. Forschungsförderungseinrichtungen in ihren Vergabekriterien; Mechanismen der Organisation wirken in Universitäten oder Forschungseinrichtungen, wenn z.B. im Generationswechsel beim Austausch von Personen neue theoretische und methodische Programme zur Geltung kommen und alte ihre Anerkennung und institutionelle Repräsentanz verlieren; schließlich werden Leistungsanreize genutzt, monetäre oder andere Gratifikationen oder der Entzug von Geld und das „Aushungern“ schwächer Programme, womit erneut das Matthäus-Prinzip<sup>4</sup> in den Wissenschaften zur Geltung kommt.

Über die Wirkung all dieser Mechanismen gibt es für die Erziehungswissenschaft bisher kaum Empirie. Die Disziplin hat sich in den letzten 30 Jahren verändert, sicherlich, aber die Gründe für diesen Wandel sind kaum präzise analysiert und Steuerungspraktiken kann man die Veränderung nur schwer zuschreiben. Man kann deshalb auch über die Wirkung mancher aktuell nahe liegender Mechanismen nur spekulieren. In der Logik der Standardindikatoren und auch in den Mechanismen disziplinärer Reproduktion liegen alte und bekannte Konsequenzen: Publikationsmaße – Frequenzen wie impact-Faktoren – könnten den Versuch nahe legen, sich dem publish or perish zu beugen; verfeinerte Bewertungsmaßstäbe formulieren die klare Erwartung, in Englisch und in ausländischen Journalen zu publizieren, nach Möglichkeit in den angesehenen und referierten Organen (die im deutschen Sprachgebiet noch immer rar sind, zum Schaden der disziplinären Selbstkontrolle). Unklar in der Publikationsdimension ist bis heute, ob

3 Bernhard Schlink sieht deshalb in aller Evaluation auch verfassungsrechtliche Probleme (Schlink 1999).

4 R.K. Merton hat das gezeigt und nach Matthäus 25, 29 benannt: „Denn wer da hat, dem wird gegeben werden, und er wird die Fülle haben; wer aber nicht hat, dem wird auch, was er hat, genommen werden.“

Funktionsrollen äquivalent wirken oder hinreichend Anerkennung sichern, also z.B. die Edition von Zeitschriften oder – bei Drittmitteln – die Tätigkeit als Gutachter (meist kommen solche belastenden Ehren aber so spät, dass sie karrieremäßig kaum mehr relevant werden). Drittmittleinwerbungen als Indikator favorisieren wiederum ein Verhalten, in dem der Kampf um Ressourcen zum Alltag wird, und zwar in gestufter Wertigkeit: Reisekosten zu Kongressen befördern die Sichtbarkeit, als forschungsrelevant gelten eher eigene Projekte, nach Möglichkeit bei den angesehenen Drittmittelgebern, noch reputierlicher ist die Gründung von Forschergruppen, steigerbar durch die Gründung oder Mitwirkung in Kollegs oder Sonderforschungsbereichen sowie, für den eigenen Nachwuchs, die Platzierung in speziellen Programmen oder die Vermittlung auf selbstständige Professuren – Schulbildung also (wenn auch mit eigenen Effekten).

Man sieht heute das Bündel an denkbaren Strategien in der Reaktion auf Evaluation, offen ist allenfalls, ob sie neben der Beförderung der Karriere auch wirklich den Status der Disziplin verändert haben oder verändern werden. Schon deshalb ist kontinuierliche Selbstbeobachtung im Fach eine dringliche und kontinuierlich zu leistende Aufgabe. Schon deshalb besteht auch Bedarf an Kommunikation über die Implikationen und Konsequenzen von Evaluation. Mit Blick auf die wiederkehrenden Krisendiagnosen über die Erziehungswissenschaft ist es wohl das zentrale ungelöste Problem, wie aus den Befunden von Selbst- und Fremdbeobachtung wirklich eine Veränderung der Disziplin zum Besseren folgen kann.

Rezepte sind an dieser Stelle schon wegen des systematischen Konflikts von Freiheit und Steuerung nicht zu erwarten, aber man darf auf die vielleicht nicht intendierten Folgen ökonomischer Rahmenbedingungen hoffen: Geldknappheit wird dazu führen, dass in Universitäten und Forschungseinrichtungen die Selbst- und Fremdkontrolle zunehmen und Evaluation damit alltäglich werden wird, alltäglich damit auch die Selbstwahrnehmung der eigenen wissenschaftlichen Praxis und ihre Dokumentation vor dem Hintergrund der Standardindikatoren. Gleichzeitig darf man erwarten, dass, der Not geschuldet und wegen der Konkurrenz zu anderen Disziplinen innerhalb der Bildungsforschung, die Bereitschaft steigen wird, sich extern mit Ressourcen zu versorgen und damit auch die externe Begutachtung der eigenen Projekte von der Antragstellung bis zur Publikation in Kauf zu nehmen.

Innerhalb der Disziplin werden dann Debatten über die Prioritäten der Forschung, über Desiderata und neue Themen zwar auch noch unmittelbar geführt, entschieden werden sie wahrscheinlich aber nicht allein wissenschaftstheoretisch, sondern eher indirekt: über die Präferenzsysteme, die z.B. in Vergabekommissionen etabliert werden, und in den bildungspolitischen Entscheidungsprozessen, mit denen Disziplinen nicht nur ihre Reputation im Konzert der als relevant beurteilten Fächer, sondern auch ihren universitären Status behalten oder verlieren, etwa im Kontext der Lehrerbildung oder der Schulreform. Am Ende werden wir feststellen, dass die Debatte über Evaluation und die Aufregung über die „Kommissare“ nur Symptome eines Wandels sind, in dem aus der guten alten Pädagogik entweder eine Erziehungswissenschaft wird, die forschungsfähig ist wie alle, oder nur eine wissenschaftliche Gemeinschaft zurückbleibt, die ihre eigene Transformation übersieht, indem sie den Wandel beklagt.

## Literatur

- Badley, G. (2003): The Crisis in educational research: a pragmatic approach. In: *European Educational Research Journal* 2, S. 296-308.
- Baumert, J./Roeder, P.M. (1990): Forschungsproduktivität und ihre institutionellen Bedingungen – Alltag erziehungswissenschaftlicher Forschung. In: *Zeitschrift für Pädagogik* 36, S. 73-98.
- Brezinka, W. (2000): *Pädagogik in Österreich. Bd.1: Einleitung. Pädagogik an der Universität Wien.* Wien (Bundesverlag).
- Brezinka, W. (2003): Krise und Zukunft der Pädagogik als Universitätsfach. In: Ders.: *Erziehung und Pädagogik im Kulturwandel.* München: Reinhardt, S. 168-185.
- Hörmann, G. (2002): Wo bleibt die Evaluation der Evaluateure (Amateure)? In: *PÄD Forum*, S. 352-361.
- Hofstetter, R./Schneuwly, B. (Eds.) (1998): *Le pari des sciences de l'éducation. (De Boeck & Larcier) Raison educativas 1.*
- Hofstetter, R./Schneuwly, B. (2001): *Educational Science in Switzerland. Evolution and Outlooks.* Bern: CEST.
- Hofstetter, R./Schneuwly, B. (2001): *Sciences de l'éducation 19e-20e siecles. Entre champs professionnels et champs disciplinaires. Erziehungswissenschaft(en) 19-20. Jahrhundert. Zwischen Profession und Disziplin.* Bern u.a.: Lang.
- Hornbostel, S./Keiner, E. (2002): Evaluation der Erziehungswissenschaft. In: *Zeitschrift für Erziehungswissenschaft* 5, S. 634-653.
- Keiner, E. (Hrsg.) (2001): *Evaluation (in) der Erziehungswissenschaft.* Weinheim, Basel: Beltz.
- Keiner, E./Schriewer, J. (2000): Erneuerung aus dem Geist der eigenen Tradition? Über Kontinuität und Wandel nationaler Denkstile in der Erziehungswissenschaft. In: *Revue suisse des sciences de l'éducation* 2, S. 27-50.
- Lagemann, E.C. (2000): *An elusive science. The troubling history of education research.* Chicago/London: University Press.
- Macke, G. (1990): Disziplinenformierung als Differenzierung und Spezialisierung. In: *Zeitschrift für Pädagogik* 36, S. 51-72.
- Martin, B.R. (2003): Research assessment in the United Kingdom and how it might be improved. In: *Berlin-Brandenburgische Akademie der Wissenschaften. Berichte und Abhandlungen, Bd. 9,* Berlin: Akademie Verlag, S. 255-266.
- Merkens, H./Dreyer, J. (2002): Professuren im Fach Erziehungswissenschaft. In: Merckens, H./Rauschenbach, T./Weishaupt, H. (Hrsg.): *Datenreport Erziehungswissenschaft 2. Ergänzende Analysen.* Opladen: Leske & Budrich, S. 125-142.
- Merkens, H./Tippelt, R. (2002): Das andere Gesicht der Erziehungswissenschaft. In: *Süddeutsche Zeitung* vom 22.10.2002.
- Merkens, H./Weishaupt, H./Zedler, P. (2000): Lokale Profile. In: Merckens, H./Rauschenbach, T./Weishaupt, H. (Hrsg.): *Datenreport Erziehungswissenschaft.* Opladen: Leske & Budrich, S. 145-154.
- Nießen, M. (2001): Die Erziehungswissenschaft im Spiegel der Forschungsförderung durch die DFG – Beobachtungen beim Blick über die Schulter. In: Keiner, E./Pollak, G. (Hrsg.): *Erziehungswissenschaft: Wissenschaftstheorie und Wissenschaftspolitik.* Weinheim, Basel: Beltz, S. 259-266.
- OECD (2002): *Educational Research and Development in England: examiner's report.* In: *European Educational Research Journal* 1, S. 705-738.
- Prange, K. (2003): [Rezension von:] H.U. Otto u.a.: *Erziehungswissenschaft in Studium und Beruf.* In: *Zeitschrift für Pädagogik* 49, S. 756-759.
- Prange, K. (i. Druck): Über die Kunst des Rezensierens. In: *Zeitschrift für Pädagogik* 50.
- Prondczynsky, A.v. (2001): Evaluation der Lehrerbildung in den USA: Geschichte, Methoden, Befunde. In: Keiner, E. (Hrsg.): *Evaluation (in) der Erziehungswissenschaft.* Weinheim, Basel: Beltz, S. 91-140.

- Rang, B. (2001): Large Scale Assessment in der Erziehungswissenschaft: Die landesweite Evaluation der niederländischen Erziehungswissenschaft 1997/98. In: Keiner, E. (Hrsg.): Evaluation (in) der Erziehungswissenschaft. Weinheim, Basel: Beltz, S. 141-154.
- Schlink, B. (1999): Öffentliche Vorlesungen, Berlin (Humboldt-Universität) 100.
- Tenorth, H.-E. (im Druck): Erziehungswissenschaft. In: Benner, D./Oelkers, J. (Hrsg.): Historisches Wörterbuch der Pädagogik. Weinheim, Basel, Berlin: Beltz.
- Van der Meulen, B. (2003): Evaluation and the governance of academic research. Experiences and recent developments in the Netherlands. In: Berlin-Brandenburgische Akademie der Wissenschaften. Berichte und Abhandlungen, Bd. 9. Berlin: Akademie Verlag, S. 267-281.
- Weiler, H. (2003): Bildungsforschung und Bildungsreform: Von den Defiziten der westdeutschen Erziehungswissenschaft. In: Gogolin, I./Tippelt, R. (Hrsg.): Innovation durch Bildung. Beiträge zum 18. Kongress der DGfE (München 2002). Opladen: Leske & Budrich, S. 181-203.
- Weishaupt, H. (2002): Aktuelle Veränderungen der institutionellen Struktur des erziehungswissenschaftlichen Forschungsfeldes. In: Merkens, H./Rauschenbach, T./Weishaupt, H. (Hrsg.): Datenreport Erziehungswissenschaft 2. Ergänzende Analysen. Opladen: Leske & Budrich, S. 143-163.
- Weishaupt, H./Merkens, H. (2000): Forschung und wissenschaftlicher Nachwuchs. In: Merkens, H./Rauschenbach, T./Weishaupt, H. (Hrsg.): Datenreport Erziehungswissenschaft. Opladen: Leske & Budrich, S. 117-134.
- Weishaupt, H./Preuschoff, C. (2002): Die Bewertung der erziehungswissenschaftlichen Forschung durch das CHE. In: Erziehungswissenschaft 25, S. 6-17.
- Wissenschaftliche Kommission Niedersachsen (Hrsg.) (2002): Berufswissenschaften der Lehrerbildung. Forschungsevaluation an niedersächsischen Hochschulen und Forschungseinrichtungen. Hannover: Wissenschaftliche Kommission.

**Abstract:** *The central question of the article is whether evaluation can contribute to the development and the improvement of research in educational science. Starting from the assumption of a link between diagnosis and construction, the author states a number of deficits in the present practice of evaluation. In general, however, he concedes that the methods and indicators usually applied in the evaluation of educational science reveal meaningful information, yet not without pointing to the need for special specifications typical of the respective disciplines, – such as, for instance, teacher training. But so far, with regard to educational science, it has not been empirically substantiated that evaluations are not only informative but also have a stimulating effect in offering certain practical conclusions.*

*Anschrift des Autors:*

Prof. Dr. Heinz-Elmar Tenorth, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin.